

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平6-28392

(43)公開日 平成6年(1994)2月4日

(51)Int.Cl. <sup>5</sup> G 0 6 F 15/38 15/18 15/38	識別記号 E 9194-5L 8945-5L J 9194-5L	序内整理番号 F I	技術表示箇所
---	---	---------------	--------

審査請求 未請求 請求項の数4(全20頁)

(21)出願番号 特願平3-43661  
(22)出願日 平成3年(1991)3月8日

(71)出願人 000005223  
富士通株式会社  
神奈川県川崎市中原区上小田中1015番地  
(72)発明者 山口 由紀子  
神奈川県川崎市中原区上小田中1015番地  
富士通株式会社内  
(74)代理人 弁理士 井島 藤治 (外1名)

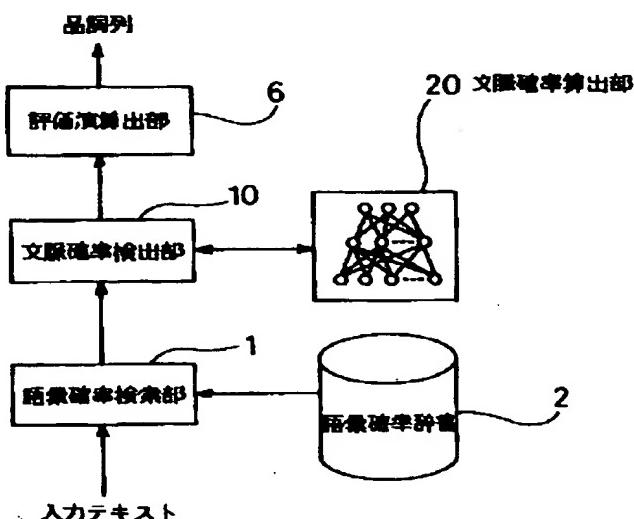
(54)【発明の名称】品詞選択システム

(57)【要約】

【目的】本発明は文章(テキスト)中の各単語の品詞の種類を自動的に選択するようにした品詞選択システムに関し、メモリ容量を小さくすることができる品詞選択システムを提供することを目的としている。

【構成】入力テキストを受けて、単語毎の語彙確率を検索する語彙確率検索部1と、語彙と該語彙の確率が格納され、前記語彙検索部1の検索の対象となる語彙確率辞書2と、前記語彙確率検索部1から検索された語彙確率を基に、複数個の品詞並びの文脈確率を検出する文脈確率検出部10と、該文脈確率検出部10から与えられる入力パターンを入力して次の単語の品詞の種類に応じた文脈確率を算出する、ニューラルネットワークを用いた文脈確率算出部20と、前記文脈確率検出部10の出力を受けて、複数個の品詞並びの文脈確率と該品詞並びの先頭または末尾の単語の語彙確率から所定の手順に従って評価値を算出する評価値算出部6から構成される。

本発明の原理ブロック図



1

## 【特許請求の範囲】

【請求項1】 入力テキストを受けて、単語毎の語彙確率を検索する語彙確率検索部(1)と、語彙と該語彙の確率が格納され、前記語彙検索部(1)の検索の対象となる語彙確率辞書(2)と、

前記語彙確率検索部(1)から検索された語彙確率を基に、複数個の品詞並びの文脈確率を検出する文脈確率検出部(10)と、

該文脈確率検出部(10)から与えられる入力パターンを入力して次の単語の品詞の種類に応じた文脈確率を算出する、ニューラルネットワークを用いた文脈確率算出部(20)と、

前記文脈確率検出部(10)の出力を受けて、複数個の品詞並びの文脈確率と該品詞並びの先頭または末尾の単語の語彙確率から所定の手順に従って評価値を算出する評価値算出部(6)から構成された品詞選択システム。

【請求項2】 前記文脈確率算出部(20)がn単語品詞並びの文脈確率を算出する場合において、

該品詞並びの先頭から(n-1)番目までの単語の品詞パターンを入力として(n-1)単語の品詞並びに対するn番目の単語の各品詞カテゴリがとる確率を出力するように構成したことを特徴とする請求項1記載の品詞選択システム。

【請求項3】 前記文脈確率算出部(20)がn単語品詞並びの文脈確率を算出する場合において、

該品詞並びの第2番目から末尾nまでの単語の品詞パターンを入力として(n-1)単語の品詞並びに対する先頭の単語の各品詞カテゴリがとる確率を出力するように構成したことを特徴とする請求項1記載の品詞選択システム。

【請求項4】 前記文脈確率算出部(20)がn単語品詞並びの文脈確率を算出する場合において、

(n-1)個の品詞並びパターンにそれぞれの品詞をとる確率を示す評価値を入力して第n番目又は先頭の単語の各品詞カテゴリがとる確率を出力するように構成したことを特徴とする請求項2又は3記載の品詞選択システム。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 本発明は自然言語処理技術に関し、更に詳しくは文章(テキスト)中の各単語の品詞の種類を自動的に選択するようにした品詞選択システムに関する。

【0002】 近年、音声合成の研究が進み、読み上げシステム、応答システム等のさまざまな分野で利用され始めている。合成音声の自然性を向上させるためには、出力する文章の言語情報を抽出し、それに応じたアクセントやイントネーションを生成するパターンを生成する必要がある。そのような言語情報の基礎をなすのが品詞である。品詞とは、名詞、動詞、副詞等の単語の種類をい

2

う。音声合成システムは、携帯性(ポータビリティ)が要求されていることから、文章の中の単語の品詞を簡単に抽出できるシステムが強く望まれている。

## 【0003】

【従来の技術】 最近、文章の中の単語の品詞を統計的な情報をを利用して決定する方式が用いられてきている。これは、例えば "A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text" (ACL proceedings 2nd Applied Natural Language Processing) に示されているように、与えられた単語がある品詞をとる確率である語彙確率と品詞並びの出現頻度を示す文脈確率を用いて文章中の各単語の品詞を決定するものである。

【0004】 図19は従来装置の構成例を示すブロック図である。図において、1は入力テキストを受けて、単語毎の語彙確率を検索する語彙確率検索部、2は語彙と該語彙の確率が格納され、前記語彙検索部1の検索の対象となる語彙確率辞書、3は語彙確率検索部1の出力を基に、各単語の品詞並びを作成する品詞並び作成部である。

【0005】 4は前記語彙確率検索部1から検索された語彙確率を基に、複数個の品詞並びの文脈確率を検出する文脈確率検出部、5は品詞並びの文脈確率を格納している文脈確率テーブル、6は前記文脈確率検出部4の出力を受けて、複数個の品詞並びの文脈確率と該品詞並びの先頭または末尾の単語の語彙確率から評価値を算出する評価値算出部である。このように構成された装置の動作を説明すれば、以下のとおりである。

【0006】 先ず、語彙確率検索部1は入力テキスト(文章)を入力し、各単語毎の語彙確率を語彙確率辞書2から読み込む。例えば、入力テキストが  
I love her very much.  
という文章であったものとする。ここでは、単語の品詞並びのカテゴリとして3単語品詞並びを考える。語彙確率検索部1は、この文章を構成している各単語の語彙確率を語彙確率辞書2から抽出する。

【0007】 語彙確率辞書2は、例えば図20に示すような構成となっている。図に示す内容は、入力された文章の単語のみを抽出したものであり、実際にはアルファベット順に全ての単語が品詞の種類毎の出現確率として格納されている。

【0008】 ここでは、文章の後から処理していく場合を考える。

【0009】 品詞並び作成部3は、先ず、文章のピリオド ". ##" に対する (Z##) という品詞並びを作成し、文脈確率検出部4に与える。ここで、#は文末または文頭を示す記号である。該文脈確率検出部4は、文脈確率テーブル5から (Z##) なる品詞並びの文脈確率

40

50

3

を抽出していく。文脈確率テーブル5には3個の品詞の組み合わせになるあらゆるパターンについての出現確率が文脈確率として格納されている。

【0010】図21は文脈確率テーブルの構成例を示す図である。図ではその一部のみを示している。また、ここでは品詞並びとして英語の場合、最も分類しやすい3単語品詞並びを用いている。それぞれの品詞並びの組み合わせがとり得る確率は、大量の文章を参考にして、そのとり得る確率を統計的に求めたものである。

【0011】評価値算出部6は、(Z##)なる品詞並びの文脈確率と(. Z)の語彙確率とを乗算して評価値とする。例えば、(Z##)の文脈確率が0.983、(. Z)の語彙確率が1.00であれば、その評価値は0.983となる。

【0012】次に、品詞並び作成部3は、単語並び“much. #”に対する考え方の組合せの品詞並び(ADJ Z #), (NOUN Z #), (ADV Z #)を作成する。文脈確率検出部4は、これら品詞並び(ADJ Z #), (NOUN Z #), (ADV Z #)に対する文脈確率を文脈確率テーブル5から抽出する。ここで、ADJは形容詞、NOUNは名詞、ADVは副詞である。図24に品詞の分類を示す。英語の場合、図に示すように、品詞の種類には20通りあり、ここではそれぞれの品詞について、図のような品詞記号を用いるものとする。

# # PPRON V PRON

となる。

[0017]

【発明が解決しようとする課題】前述した従来装置の場合、品詞並びの文脈確率をテーブル（文脈確率テーブル5）で保存しているため、品詞分類のカテゴリ数や品詞並びの数が大きくなるにつれて、大量のメモリが必要になるという問題があった。

【0018】本発明はこのような課題に鑑みてなされたものであって、メモリ容量を小さくすることができる品詞選択システムを提供することを目的としている。

{0019}

【課題を解決するための手段】図1は本発明の原理ブロック図である。図19と同一のものは、同一の符号をして示す。図において、1は入力テキストを受けて、単語毎の語彙確率を検索する語彙確率検索部、2は語彙と該語彙の確率が格納され、前記語彙検索部1の検索の対象となる語彙確率辞書である。

【0020】10は前記語彙確率検索部1から検索された語彙確率を基に、複数個の品詞並びの文脈確率を検出する文脈確率検出部、20は該文脈確率検出部10から与えられる入力パターンを入力して次の単語の品詞の種類に応じた文脈確率を算出する、ニューラルネットワークを用いた文脈確率算出部、6は前記文脈確率検出部10の出力を受けて、複数個の品詞並びの文脈確率と該品詞

4

\* 【0013】評価値算出部6は、それぞれの品詞並びの場合において、“much”がADJをとる場合、NO UNをとる場合及びADVをとる場合の文脈確率と語彙確率とそれまでの対応する累積評価値を乗算して評価値を算出する。

【0014】次に、品詞並び作成部3は単語並び“very much.”に対する品詞並び(ADV ADJ Z), (ADV NOUNZ Z), (ADV ADVZ), (ADJ ADJ Z), (ADJ NOUNZ Z), (ADJ ADV Z)を作成する。

【0015】文脈確率検出部4は、これら品詞並び（ADV ADJ Z), (ADV NOUNZ Z), (ADV ADV Z), (ADJ ADJ Z), (ADJNOUNZ Z), (ADJ ADV Z)に対する文脈確率を文脈確率テーブル5から抽出する。評価値算出部6は、それぞれの品詞並びの場合において、“very”がADVをとる場合及びADJをとる場合のそれぞれについて、文脈確率、語彙確率、累積評価値を乗算してそれぞれの品詞並びの評価値を算出する。

【0016】以下、同様の操作を“##1”まで繰り返すと、5個の品詞並び

love her very much.

とりうる全ての品詞並びに対する評価値が求まる。そして、それら評価値の最も高いものを品詞列として選択する。この場合に、選択される品詞列は、

**ADV ADV Z # #**

※ 詞並びの先頭または末尾の単語の語彙確率から所定の手順に従って評価値を算出する評価値算出部である。

0021

**【作用】** 予め、文脈確率算出部20に品詞並びの確率を学習させておく。例えば、3単語品詞並びの場合には、ニューラルネットワークに後ろから2つの品詞の入力パターンを入力させ、その時の次の単語（先頭単語）の品詞の種類に応じた確率を予め教師パターンとして与えてやり、教師パターンとニューラルネットワークの出力とが等しくなるように学習させておく。

【0022】しかる後、文脈確率算出部20は、文脈確率検出部10から与えられる品詞並びのパターンを入力して、次の単語の品詞の種類に応じた文脈確率を算出する。例えば、3単語品詞並びの場合には、2個の品詞並びを入力パターンとして入力し、次の単語の品詞並びをとる確率（文脈確率）が品詞の種類毎に算出される。この算出された文脈確率を基に、評価値算出部6は3単語品詞並びのパターン毎の評価値を算出する。

【0023】このように、本発明によれば文脈確率算出部20を構成するニューラルネットワークに品詞並びの確率を学習させることにより、例えば3品詞並びの場合には、後ろから2つの品詞に相当するパターンを入力パターンとして入力してやれば、次の3つ目の品詞（先頭単語の品詞）がとる確率（文脈確率）を品詞の種類ごと

5

に出力するので、文脈確率の情報を文脈確率テーブルとしてもっている必要がなくなり、メモリ容量を小さくすることができる品詞選択システムを提供することができる。

## 【0024】

【実施例】以下、図面を参照して本発明の動作を詳細に説明する。

【0025】図2は文脈確率算出部20の構成例を示す図である。21はニューラルネットワークで、入力層L1、中間層L2及び出力層L3から構成されている。入力層L1は品詞カテゴリ数の組合せが（品詞並びの数-1）個設けられている。例えば品詞並びが前述した3個である場合、入力層L1は品詞カテゴリ数の組合せが2個で構成される。出力装置L3は1個のカテゴリ数の組合せをもっている。これら品詞カテゴリ数の組合せの単位としては、例えば英語の品詞の数である20個（20ユニット）が用いられる。

【0026】22は出力層L3の出力値と教師パターンとを比較する比較部である。該比較部22は、出力層L3の出力（確率）が教師パターンと等しくなるように入力層L1と中間層L2間のシナプスの重みの変更及び中間層L2と出力層L3間の重みの変更を行う。最終的に、出力層L3の出力値と教師パターンが一致した時点で学習が終了したことになる。この時、入力層L1と中間層L2間のシナプスの重み及び中間層L2と出力層L3間の重みはある一定値に固定される。

【0027】図3は本発明の一実施例を示す構成ブロック図である。図1と同一のものは、同一の符号を伏して示す。図において、文脈確率検出部10は、語彙確率検索部1からの単語を入力して所定の数（例えば3個）だけの品詞並びを作成する品詞並び作成部11、品詞並び作成部11の出力を受けて入力パターンを作成する入力パターン作成部12、該入力パターン作成部12の出力を文脈確率算出部20に入力してニューラルネットワークから得られたそれぞれの品詞毎の文脈確率を受けて出力値を選択する出力値選択部13から構成されている。

【0028】評価値算出部6は、出力値選択部13から与えられた文脈確率と語彙確率とから評価値を算出する評価値算出手段31と、該評価値算出手段31からの複数の出力のうちから評価値が最大のものを選択する最大値選択部32から構成されている。そして、最大値選択部32の出力が求める品詞列となる。このように構成された装置の動作を説明すれば、以下のとおりである。

【0029】この実施例におけるニューラルネットワーク21は、予め大量のデータから抽出した品詞並びの頻度から、入力テキストについて末尾の単語から先頭の単語に向かって逆向きに品詞を選択して学習させる方法と、先頭の単語から末期の単語に向かって前向きに品詞を選択して学習せざる方法のいずれも用いることができる。

6

（末尾の単語から品詞を選択する場合）図4は末尾の単語から品詞を選択する場合の学習パターンの作成例を示す図である。ここでは3単語品詞並びを用いるものとする。例えば、

ADV ADV ADV

のように副詞が3個並ぶ場合を考える。このような品詞並びの頻度は大量のデータを検索した時、307個であり、その文脈確率は0.07である。この0.07は、頻度307を3単語品詞並びの後ろ2品詞が同じものの頻度の総和で割った値として求められる。このようにして、次の品詞並び

NOT ADV ADV

のようなNOTと2個の副詞の並びの頻度は78であり、その文脈確率は0.02である。以下同様にして、後の2個が全てADVで先頭の1個がカテゴリ数20のあらゆる品詞をとる確率が文脈確率として図4に示されている。この文脈確率をニューラルネットワーク21に学習させる方法について説明する。

【0030】ここでは、図2のニューラルネットワーク21が入力は2組であり、その品詞種別はいずれもADVであるものとする。このADVを示すのは、20ユニットのうちの最初のユニットに“1”を立てた状態で示す。ここで、品詞カテゴリ数として20ユニットを用いたのは、英語の場合、品詞の種類が図24に示すように20種類であることに対応させたものである。

【0031】図5(a)はこの時の入力パターン例を示す図である。今、入力パターン作成部12は図5に示すようなパターンをニューラルネットワーク21に入力させる。また、例えば、3単語品詞並びとして後の2個がNOUN ADVである場合には、その入力パターンは図5(b)に示すようなものとなる。NOUNの場合、その番号は図24より“4”である。図24は0から番号が始まっているので、“4”であることは第5番目を示す。そこで、NOUNの場合には前から5番目に“1”が立っている。

【0032】図6はこの時の学習の説明図である。入力層L1は20ユニットずつ2組で合計40ユニットの入力を受ける。各組の最初のユニットに“1”が立っているのは副詞であることを示している。出力は20ユニットである。一方、このニューラルネットワーク21の比較部22に与える教師パターンは、図7に示すようなものである。この教師パターンは図4に示す文脈確率そのまま利用したものである。

【0033】つまり、ADV ADV ADVとなる品詞並びの確率は0.07であるので、この値を教師信号として出力層L3の20ユニットの最初のユニットに与える。次のユニットには

NOT ADV ADV

となる品詞並びの確率である0.02を与える。以下、図7の値を教師パターンとして比較部22に与えてい

50

7

く。最後のユニットには

# ADV ADV

となる品詞並びの確率である0.08を与える。図6の構成において、出力層L3の各ユニットの出力が0.07, 0.02, …, 0.08をとるようになるまで何回も同一の入力パターンを与えてニューラルネットワーク21の学習を行なわせる。

【0034】以上、後の2個がADV, ADVの場合の先頭の単語がとる品詞の種類に応じた文脈確率を求める場合を例にとったが、品詞並びはこれに限るものではなく、図5(b)に示したものの他、極めて多くのパターン(例えば20カテゴリの場合3品詞並びは8000、学習パターンは400)が存在する。それぞれのパターンの組み合わせに対しても、図6のニューラルネットワーク21に学習を行なわせる。

【0035】このようにして、全ての3単語品詞並びについての学習が終了した状態で、図3に示す装置に入力テキストを入力してやると、文脈確率算出部20は入力パターンに応じた3個の品詞並びの文脈確率を出力する。この文脈確率算出部の出力から目的の品詞並びの確率の出力値を出力値選択部13が選択して、評価値算出手段31に与える。

【0036】評価値算出手段31は前記文脈確率と先頭単語の語彙確率と、それまでの対応する品詞列の累積評価値を乗算して評価値を算出する。評価値算出手段31は3品詞並びがとり得る全ての場合についての評価値を算出するので、最大値選択部32はこれら評価値の中で一番大きい値の品詞並びを選択し、求めるべき品詞列として出力する。

(先頭の単語から品詞を選択する場合)先頭の単語から品詞を選択する場合も、前述した末尾の単語から品詞を選択する場合と同様に考えることができる。

【0037】図8は末尾の単語から品詞を選択する場合の学習パターンの作成例を示す図である。ここでは3単語品詞並びを用いるものとする。例えば、

ADV ADV ADV

のように副詞が3個並ぶ場合を考える。このような品詞並びの頻度は大量のデータを検索した時、307個であり、その文脈確率は0.07である。この0.07は、頻度307を3単語品詞並びの前2品詞がADV, ADVである3品詞並びの頻度の総和で割った値として求められる。このようにして、次の品詞並び

ADV ADV NOT

のようなNOTと2個の副詞の並びの頻度は6であり、その文脈確率は0.00である。以下同様にして、前の2個が全てADVで末尾の1個がカテゴリ数20のあらゆる品詞をとる確率が文脈確率として図8に示されている。この文脈確率をニューラルネットワーク21に学習させる方法について説明する。

【0038】ここでは、入力は2組であり、その品詞種

8

別はいずれもADVであるものとする。このADVを示すのは、20ユニットのうちの最初のユニットに“1”を立てた状態で示す。

【0039】図9はこの時の入力パターン例を示す図である。今、入力パターン作成部12は図9に示すようなパターンをニューラルネットワーク21に入力させる。この時のニューラルネットワークの構成は図6と同じである。

【0040】入力層L1は20ユニットずつ2組で合計40ユニットの入力を受ける。各組の最初のユニットに“1”が立っているのは副詞であることを示している。出力は20ユニットである。一方、このニューラルネットワーク21の比較部22に与える教師パターンは、図10に示すようなものである。この教師パターンは図8に示す文脈確率をそのまま利用したものである。

【0041】つまり、ADV ADV ADVとなる品詞並びの確率は0.07であるので、この値を教師信号として出力層L3の20ユニットの最初のユニットに与える。次のユニットには

ADV ADV NOT

となる品詞並びの確率である0.00を与える。以下、図10の値を教師パターンとして比較部22に与えていく。最後のユニットには

ADV ADV #

となる品詞並びの確率である0.00を与える。図6の構成において、出力層L3の各ユニットの出力が0.07, 0.00, …, 0.00をとるようになるまで何回も同一の入力パターンを与えてニューラルネットワーク21の学習を行なわせる。

【0042】以上、前から2個がADV, ADVの場合の末尾の単語がとる品詞の種類に応じた文脈確率を求める場合を例にとったが、品詞並びはこれに限るものではなく、極めて多くのパターンが存在する。それぞれのパターンの組み合わせに対しても、図6のニューラルネットワーク21に学習を行なわせる。

【0043】このようにして、全ての3単語品詞並びについての学習が終了した状態で、図3に示す装置に入力テキストを入力してやると、文脈確率算出部20は入力パターンに応じた3個の品詞並びの文脈確率を出力する。この文脈確率算出部の出力から目的の品詞並びの確率の出力値を出力値選択部13が選択して、評価値算出手段31に与える。

【0044】評価値算出手段31は前記文脈確率と末尾単語の語彙確率と、それまでの対応する品詞列の累積評価値を乗算して評価値を算出する。評価値算出手段31は3品詞並びがとり得る全ての場合についての評価値を算出するので、最大値選択部32はこれら評価値の中で一番大きい値の品詞並びを選択し、求めるべき品詞列として出力する。

【0045】次に、具体例を用いて本発明装置の動作を

説明する。ここでは、図24に示した20カテゴリに分類した品詞を使用し、3品詞並びの文脈確率を利用して、入力テキストの各単語の品詞を末尾から先頭に向かって選択していく場合を考える。ここでは、

I appealed to the children to make less noise.

という文章について品詞選択の処理を行う。語彙確率検索部1は、入力テキスト中の各単語について語彙確率辞書2を検索し、それぞれの単語について品詞カテゴリとその確率を抽出する。図11はこのようにして抽出された品詞カテゴリとその確率を示す図である。前記文章に対応して、それぞれの単語の品詞カテゴリとその語彙確率とが示されている。

【0046】品詞並び作成部11は、この検索結果について3単語の品詞並びを作成する。図12はこのようにして作成された3単語品詞並びを示す図である。文章の後の方から実現される可能性のある全ての品詞並びが作成されている。

【0047】入力パターン作成部12は、これらの品詞並びの後2品詞から図5に示したような入力パターンを作成し、ニューラルネットワークを用いた文脈確率算出部20に与える。文脈確率算出部20は、後2品詞から先頭の単語のとる確率を文脈確率として品詞カテゴリ毎に出力する。

【0048】出力値選択部13は、与えられた入力パターンに対する文脈確率算出部20の出力について、品詞並びの先頭の品詞に対する値を選択し、当該品詞並びの文脈確率とする。

【0049】評価値算出手段31は、出力値選択部13から得られた文脈確率Pbと、語彙確率検索部1で検索した品詞並びの先頭の品詞に対する語彙確率Pvを、入力テキストの末尾の単語から作成した品詞並びに対する累積評価値e'にかけた値eを新しい評価値とする。新しい評価値eは次式で表される。

【0050】

$$e = Pb \times Pv \times e' \quad (1)$$

図13、図14はこのようにして得られた評価値算出手段31の結果を示す図である。図13と図14は連続しており、図13の後に図14がくるようになっている。図において、Aの部分は文脈確率算出部20で前述したシーケンスにより得られた文脈確率、Bは先頭単語の品詞の種類に応じた語彙確率、Cはそれまでの過程で得られた累積評価値e、Dは当該累積評価値に対応した品詞並びをそれぞれ示している。

【0051】入力テキストの末尾の単語から順次、当該単語がとる品詞カテゴリの全てについて、それまでに作成した品詞並びの全てについて当該品詞並びの後2品詞と合わせて3品詞並びを作成し、文脈確率Pbを検索する。この文脈確率Pbと、当該単語の品詞カテゴリにおける語彙確率Pvと、それまでに作成した品詞並びの評価値e'の積を新しい品詞並びの評価値eとしている。

図に示すように、入力テキストの先頭の単語まで、全ての品詞並びの評価値が算出できた後、最大値選択部32はこれらの品詞並びの中から最大の評価値を持つものを選択し、入力テキストの各単語の品詞を決定する。

【0052】図15はこのようにして得られた品詞選択結果を示す図である。図14の最終段の品詞並びと一致している。

【0053】図16は本発明の他の実施例を示す構成ブロック図である。図3と同一のものは、同一の符号をして示す。図において、1は入力テキストを受けて、単語毎の語彙確率を検索する語彙確率検索部、2は語彙と該語彙の確率が格納され、前記語彙検索部1の検索の対象となる語彙確率辞書、10は語彙確率検索部1の出力を受けて入力バターンを作成する文脈確率検出部としての入力バターン作成部である。

【0054】20は入力バターン作成部10から与えられる入力バターンを入力して次の単語の品詞の種類に応じた文脈確率を算出する、ニューラルネットワークを用いた文脈確率算出部、31は文脈確率算出部20の出力を受けて、与えられた文脈確率と語彙確率とから評価値を算出する評価値算出手段、32は該評価値算出手段31からの複数の出力のうちから評価値が最大のものを選択する最大値選択部32である。これら、評価値算出手段31と最大値選択部32とで評価値算出部6を構成している。このように構成された装置の動作を説明すれば、以下のとおりである。

【0055】この実施例では、後2品詞並びの入力パターンを図5に示すように“1”か“0”かで入力するのではなく、後2品詞並びを構成する単語の各品詞カテゴリをとる確率を示す評価値を入力するようにしたものである。

【0056】図3の実施例と同様に、語彙確率検索部1は入力テキスト中の各単語について語彙確率辞書2を検索して、それぞれの単語についての品詞カテゴリとその確率を図11に示すように抽出する。ここでは、品詞を選択する際の評価値を単語毎に品詞カテゴリの語彙確率で表現し、ニューラルネットワークを用いた文脈確率算出部20に対して隣接する後2単語又は前2単語の品詞カテゴリの確率を入力する。

【0057】図17、図18は評価値算出手段31の例を示す図である。図17、図18は一連の動作を示したものであり、図17の後に図18が続くようになっている。入力テキストの末尾の単語から品詞を選択する場合には、入力パターン作成部10は、3単語品詞並びのうちの後2単語の評価値Eをニューラルネットワーク21に入力し、得られた出力が文脈確率Pbとなる。

【0058】この文脈確率Pbのそれぞれの品詞カテゴリの確率に、語彙確率検索部1で得られた該単語の各品詞カテゴリの確率をかけたものを当該単語の評価値Eとする。

11

【0059】最大値選択部32は、入力テキストの各単語について、評価値Eにおいて最大の値となっている品詞カテゴリを当該単語の品詞として選択する。例えば、図17の場合において、Wn=6の場合には単語として“make”を評価している。この時の評価値Eは、前から10番目のユニットが0.34と最大値をとっている。前から10番目の品詞は、図22より動詞(V)である。このようにして、品詞を選択した結果も、図15と同じになる。

【0060】この実施例によれば、ニューラルネットワーク21に品詞カテゴリをとる確率を示す評価値を入力することにより、文脈確率の検索の回数を減少させることができとなり、高速に品詞選択を行うことができる。

【0061】上述の実施例では英語に本発明を適用した場合について説明したが、本発明はこれに限るものではない。日本語その他の文章についても本発明を適用することができる。

#### 【0062】

【発明の効果】以上、詳細に説明したように、本発明によれば文脈確率を求めるのにニューラルネットワークを用いるようにすることにより、文脈確率テーブルを持つ必要がなくなり、メモリ容量を小さくすることができる品詞選択システムを提供することができる。

#### 【図面の簡単な説明】

【図1】本発明の原理ブロック図である。

【図2】文脈確率算出部の構成例を示す図である。

【図3】本発明の一実施例を示す構成ブロック図である。

【図4】末尾の単語から品詞を選択する場合の学習パターンの作成例を示す図である。

【図7】

教師パターン例を示す図

【図10】

教師パターン例を示す図

【図21】

文脈確率テーブルの構成例を示す図

#### 教師パターン例を示す図

0.07	0.02	0.01	0.03	0.14	0.01	0.05	0.00	0.02	0.19
------	------	------	------	------	------	------	------	------	------

0.01	0.14	0.01	0.01	0.02	0.00	0.07	0.00	0.11	0.08
------	------	------	------	------	------	------	------	------	------

#### 教師パターン例を示す図

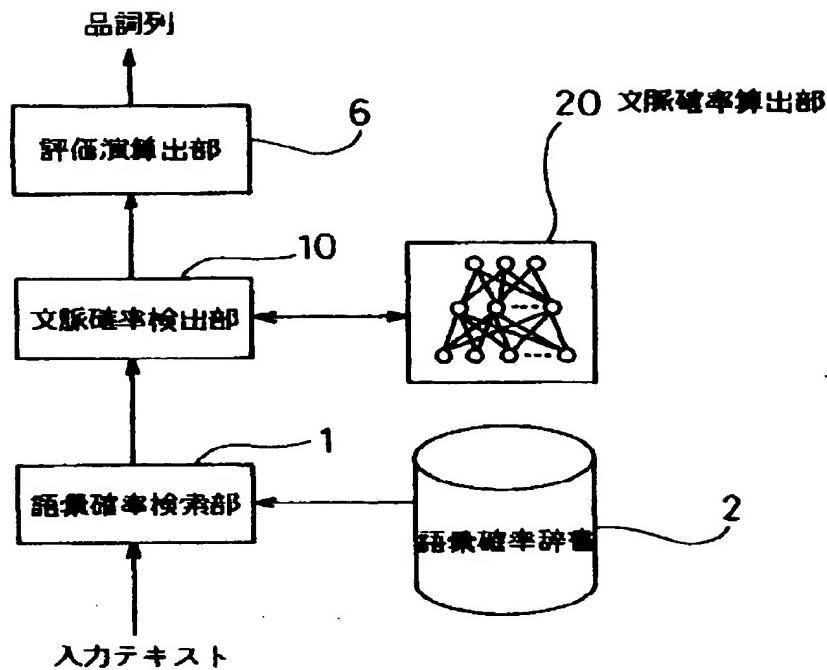
0.07	0.00	0.00	0.05	0.02	0.00	0.03	0.01	0.13	0.13
------	------	------	------	------	------	------	------	------	------

0.00	0.01	0.00	0.00	0.11	0.01	0.18	0.00	0.21	0.00
------	------	------	------	------	------	------	------	------	------

3 単語品詞並び			文脈確率
ADV	ADV	ADV	0.07
ADV	ADV	NOT	0.00
ADV	ADV	ADVP	0.00
ADV	ADV	DET	0.05
ADV	ADV	NOUN	0.02
ADV	ADV	PRON	0.00
ADV	ADV	PPRON	0.05
ADV	ADV	NUM	0.01
ADV	ADV	ADJ	0.13
ADV	ADV	VERB	0.13
ADV	ADV	AUX	0.00

【図1】

本発明の原理プロック図



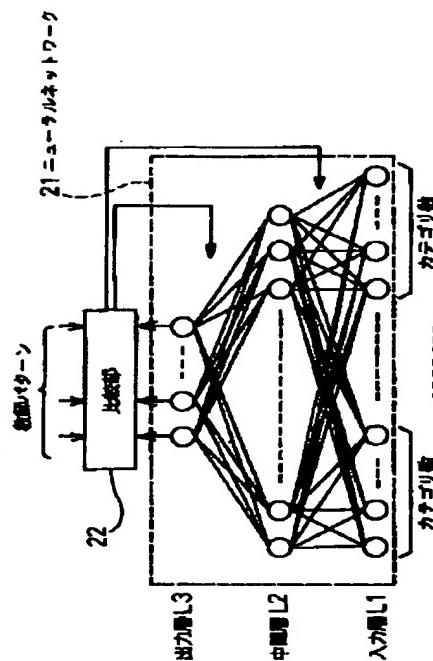
【図11】

抽出された品詞カテゴリとその確率を示す図

I		NOUN	0.00
		PPTON	1.00
appealed	appealed	VERB	1.00
to	to	ADV	0.00
	to	PREP	0.43
	to	TO	0.57
the	the	DET	1.00
children	children	NOUN	1.00
to	to	ADV	0.00
	to	PREP	0.43
	to	TO	0.57
make	make	NOUN	0.01
	make	VERB	0.99
less	less	ADV	0.47
	less	ADJ	0.53
noise	noise	NOUN	0.97
	noise	VERB	0.03
		Z	1.00

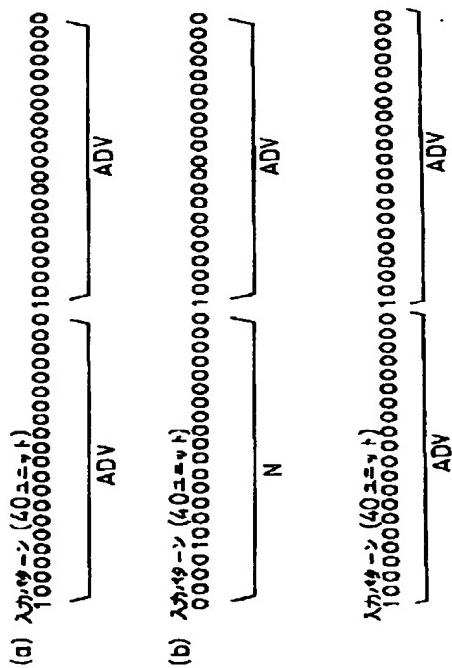
【図2】

文脈確率検出部の構成例を示す図



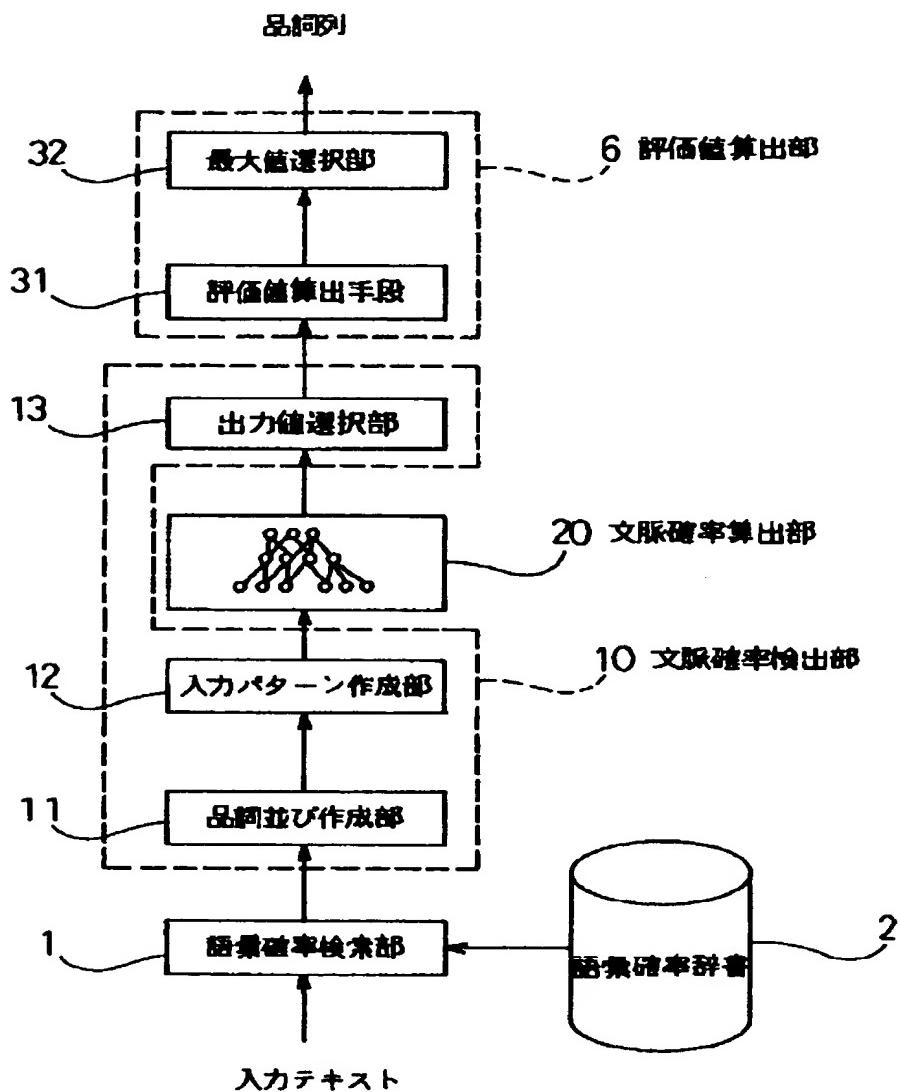
【図5】

入力パターン例を示す図



【図3】

本発明の一実施例を示す構成ブロック図



【図15】

本発明による品詞選択結果を示す図

単語	品詞
I	PRON
appealed	VERB
to	PREP
the	DET
children	NOUN
to	TO
make	VERB
less	ADJ
noise	NOUN
	Z

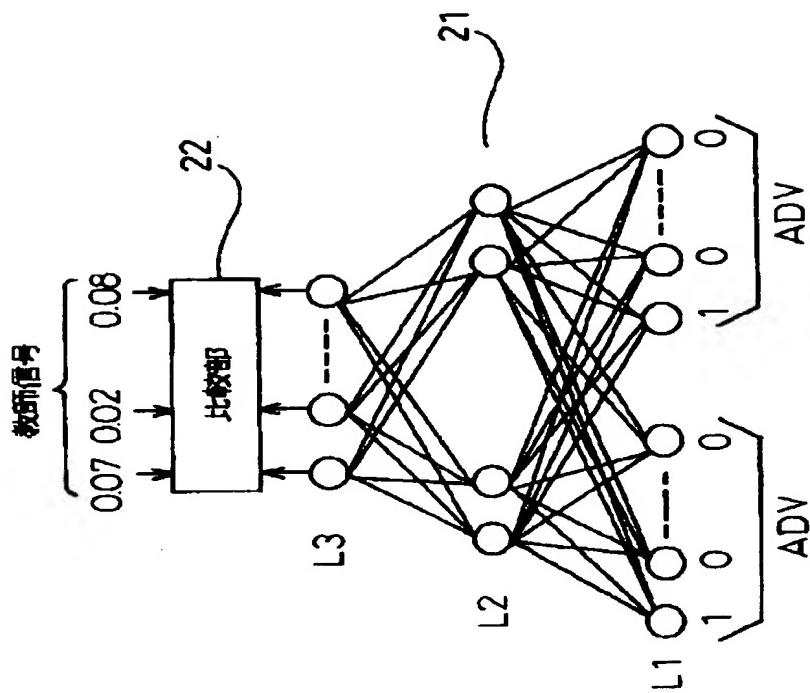
【図4】

末尾の単語から品詞を選択する場合の学習パターンの作成例を示す図

3 単語品詞並び	頻度	文脈確率
ADV ADV ADV	307	0.07
NOT ADV ADV	78	0.02
ADVP ADV ADV	53	0.01
DBT ADV ADV	139	0.03
NOUN ADV ADV	642	0.14
PRON ADV ADV	30	0.01
PPRON ADV ADV	222	0.05
NUN ADV ADV	15	0.00
ADJ ADV ADV	67	0.02
VERB ADV ADV	848	0.19
AUX ADV ADV	65	0.01
BE ADV ADV	641	0.14
DO ADV ADV	24	0.01
HAVE ADV ADV	55	0.01
PREP ADV ADV	76	0.02
TO ADV ADV	0	0.00
CJ ADV ADV	320	0.07
INTJ ADV ADV	0	0.00
Z ADV ADV	507	0.11
# ADV ADV	348	0.08

【図6】

学習の説明図



【図12】

作成された3単語の品詞並びを示す図

(Z * *)	(NOUN TO VERB)
(NOUN Z * *)	(NOUN PREP NOUN)
(VERB Z * *)	(NOUN PREP VERB)
(NOUN ADV NOUN)	(NOUN ADV NOUN)
(ADJ NOUN Z *)	(DET NOUN ADV)
(ADJ NOUN Z *)	(DET NOUN PREP)
(ADJ NOUN Z *)	(DET NOUN TO )
(NOUN ADV NOUN)	(ADV DET NOUN)
(NOUN ADV VERB)	(PREP DET NOUN)
(NOUN ADV NOUN)	(TO DET NOUN)
(VERB ADV NOUN)	(VERB TO DET )
(VERB ADV VERB)	(VERB PREP DET )
(VERB ADV NOUN)	(VERB ADV DET )
(VERB ADV VERB)	(NOUN VERB ADV )
(ADV VERB ADV )	(NOUN VERB PREP)
(ADV VERB ADJ )	(NOUN VERB TO )
(ADV NOUN ADJ )	(PRON VERB ADV )
(PREP VERB ADV )	(PRON VERB PREP)
(PREP VERB ADJ )	(PRON VERB TO )
(PREP NOUN ADV )	(NOUN VERB ADV )
(TO VERB ADV )	(NOUN VERB PREP)
(TO VERB ADJ )	(NOUN VERB TO )
(TO NOUN ADV )	(# PRON VERB )
(TO NOUN ADJ )	(# NOUN VERB )
(NOUN TO NOUN)	(# # NOUN )
	(# # PRON )

【図8】

先頭の単語から品詞を選択する場合の学習パターンの作成例を示す図

3単語品詞並び	頻度	文脈確率
ADV ADV ADV	307	0.07
ADV ADV NOT	6	0.00
ADV ADV ADVP	8	0.00
ADV ADV DET	221	0.05
ADV ADV NOUN	79	0.02
ADV ADV PRON	19	0.00
ADV ADV PPRON	216	0.05
ADV ADV NUM	27	0.01
ADV ADV ADJ	500	0.13
ADV ADV VERB	592	0.13
ADV ADV AUX	22	0.00
ADV ADV BE	57	0.01
ADV ADV DO	19	0.00
ADV ADV HAVE	15	0.00
ADV ADV PREP	466	0.11
ADV ADV TO	66	0.01
ADV ADV CJ	783	0.18
ADV ADV INTJ	0	0.00
ADV ADV Z	945	0.21
ADV ADV #	0	0.00

【図13】

本発明により得られた評価値算出結果を示す図

A		B		C		D	
(2	)	= 0.983	(1, 2) = 1.000			0 = 9.330430e-01	2.9 1
NOUN	Z'	= 0.433	(nouns NOUN) = 0.974			0 = 4.744334e-01	NOUN 2.0 1
VERB	Z'	= 0.368	(nouns VERB) = 0.026			0 = 6.751982e-03	VERB 2.0 1
ADV	VERB	= 0.260	(less ADV) = 0.469			0 = 8.237001e-04	ADV VERB 2.0 1
ADV	NOUN	= 0.142	(less ADV) = 0.469			0 = 2.754618e-02	ADV NOUN 2.0 1
ADV	VERB	= 0.021	(less ADV) = 0.531			0 = 7.682376e-05	ADV VERB 2.0 1
ADV	NOUN	= 0.205	(less ADV) = 0.531			0 = 4.598197e-02	ADV NOUN 2.0 1
NOUN	ADJ	= 0.000	(make NOUN) = 0.005			0 = 1.131469e-10	NOUN ADJ NOUN 2.0 1
NOUN	ADV	= 0.000	(make NOUN) = 0.005			0 = 1.922805e-13	NOUN ADV VERB 2.0 1
DIOSR	ADJ	= 0.010	(make NOUN) = 0.005			0 = 1.443124e-06	NOUN ADV NOUN 2.0 1
NOUN	ADV	= 0.009	(make NOUN) = 0.005			0 = 3.401198e-08	NOUN ADV VERB 2.0 1
VERB	ADJ	= 0.185	(make VERB) = 0.995			0 = 8.319324e-03	VERB ADJ NOUN 2.0 1
VERB	ADV	= 0.190	(make VERB) = 0.995			0 = 1.450862e-03	VERB ADV VERB 2.0 1
VERB	ADJ	= 0.247	(make VERB) = 0.995			0 = 6.757389e-03	VERB ADJ NOUN 2.0 1
VERB	ADV	= 0.213	(make VERB) = 0.995			0 = 1.709512e-04	VERB ADV VERB 2.0 1
ADV	VERB	= 0.243	(to ADV) = 0.000			0 = 5.333929e-08	ADV VERB ADV NOUN 2.0 1
ADV	VERB	= 0.233	(to ADV) = 0.000			0 = 7.408467e-08	ADV VERB ADV NOUN 2.0 1
ADV	NOUN	= 0.125	(to ADV) = 0.000			0 = 6.362238e-12	ADV NOUN ADV NOUN 2.0 1
ADV	NOUN	= 0.154	(to ADV) = 0.000			0 = 6.649494e-16	ADV NOUN ADV NOUN 2.0 1
PREP	VERB	= 0.055	(to PREP) = 0.426			0 = 1.582023e-04	PREP VERB ADV NOUN 2.0 1
PREP	VERB	= 0.050	(to PREP) = 0.426			0 = 1.780844e-04	PREP VERB ADV NOUN 2.0 1
PREP	NOUN	= 0.088	(to PREP) = 0.426			0 = 5.419616e-04	PREP NOUN ADV NOUN 2.0 1
PREP	NOUN	= 0.134	(to PREP) = 0.426			0 = 7.413142e-12	PREP NOUN ADV NOUN 2.0 1
TO	VERB	= 0.130	(to TO) = 0.374			0 = 5.040969e-04	TO VERB ADV NOUN 2.0 1
TO	VERB	= 0.178	(to TO) = 0.574			0 = 8.513891e-04	TO VERB ADV NOUN 2.0 1
TO	NOUN	= 0.037	(to TO) = 0.574			0 = 3.085738e-08	TO NOUN ADV NOUN 2.0 1
TO	NOUN	= 0.041	(to TO) = 0.574			0 = 2.6411142e-12	TO NOUN ADV NOUN 2.0 1
NOUN	TO	= 0.225	(children NOUN) = 1.000			0 = 6.942779e-09	NOUN TO NOUN ADV NOUN 2.0 1
NOUN	TO	= 0.103	(children NOUN) = 1.000			0 = 6.732509e-05	NOUN TO VERB ADV NOUN 2.0 1
NOUN	PREP	= 0.413	(children NOUN) = 1.000			0 = 2.2329889e-08	NOUN PREP NOUN ADV NOUN 2.0 1
NOUN	PREP	= 0.341	(children NOUN) = 1.000			0 = 6.079481e-03	NOUN PREP VERB ADV NOUN 2.0 1
NOUN	ADV	= 0.010	(children NOUN) = 1.000			0 = 7.204993e-14	NOUN ADV NOUN ADV NOUN 2.0 1
NOUN	ADV	= 0.036	(children NOUN) = 1.000			0 = 6.095321e-10	NOUN ADV VERB ADV NOUN 2.0 1

【図14】

本発明により得られた評価値算出結果を示す図

			D
A		B	C
DET	NOUN ADV ) = 0.350	(the DET) * 1.000	* = 2.130362e-10 DET NOUN ADV VERB ADJ NOUN 2 1 1
NOUN PREP )	= 0.378	(the DET) * 1.000	* = 2.288917e-05 DET NOUN PREP VERB ADJ NOUN 1 1 1
DET NOUN TO )	= 0.385	(the DET) * 1.000	* = 3.393971e-05 DET NOUN TO VERB ADJ NOUN 2 1 1
ADV DET NOUN )	= 0.160	(so ADV) * 0.000	* = 2.0560934e-10 ADV DET NOUN TO VERB ADJ NOUN 2 1 1
PREP DET NOUN )	= 0.487	(so PREP) * 0.426	* = 6.986655e-06 PREP DET NOUN TO VERB ADJ NOUN 2 1 1
PREP DET NOUN )	= 0.052	(so TO) * 0.574	* = 9.389269e-07 TO DET NOUN TO VERB ADJ NOUN 2 1 1
VERB TO DET )	= 0.000	(appeared VERB) * 1.000	* = 1.994630e-13 VERB TO DET NOUN TO VERB ADJ NOUN 2 1 1
VERB PREP DET )	= 0.424	(appeared VERB) * 1.000	* = 2.951669e-06 VERB PREP DET NOUN TO VERB ADJ NOUN 2 1 1
VERB ADV DET )	= 0.342	(appeared VERB) * 1.000	* = 7.043271e-11 VERB ADV DET NOUN TO VERB ADJ NOUN 2 1 1
NOUN VERB ADV )	= 0.032	(I NOUN) * 0.000	* = 8.854643e-16 NOUN VERB ADV DET NOUN TO VERB ADJ NOUN 2 1 1
NOUN VERB PREP )	= 0.314	(I NOUN) * 0.000	* = 1.608170e-11 NOUN VERB PREP DET NOUN TO VERB ADJ NOUN 2 1 1
NOUN VERB TO )	= 0.011	(I NOUN) * 0.000	* = 7.963567e-12 NOUN VERB TO DET NOUN TO VERB ADJ NOUN 2 1 1
IPRON VERB ADV )	= 0.155	(I PP(ON)) * 0.999	* = 1.161925e-11 PRON VERB ADV DET NOUN TO VERB ADJ NOUN 2 1 1
IPRON VERB PREP )	= 0.167	(I PP(ON)) * 0.999	* = 4.321574e-07 PRON VERB PREP DET NOUN TO VERB ADJ NOUN 2 1 1
IPRON VERB TO )	= 0.156	(I PP(ON)) * 0.999	* = 9.464936e-14 PRON VERB TO DET NOUN TO VERB ADJ NOUN 2 1 1
NOUN VERB ADV )	= 0.032	(I NOUN) * 0.000	* = 9.427210e-16 NOUN VERB ADV DET NOUN TO VERB ADJ NOUN 2 1 1
NOUN VERB PREP )	= 0.014	(I NOUN) * 0.000	* = 8.0406534e-12 NOUN VERB PREP DET NOUN TO VERB ADJ NOUN 2 1 1
NOUN VERB TO )	= 0.041	(I NOUN) * 0.000	* = 3.880793e-18 NOUN VERB TO DET NOUN TO VERB ADJ NOUN 2 1 1
IPRON VERB )	= 0.128	(I) * 1.000	* = 1.617827e-07 PRON VERB PREP DET NOUN TO VERB ADJ NOUN 2 1 1
I NOUN VERB )	= 0.116	(I) * 1.000	* = 2.150194e-12 NOUN VERB PREP DET NOUN TO VERB ADJ NOUN 2 1 1
I ) NOUN )	= 1.000	(I) * 1.000	* = 2.150194e-12 NOUN VERB PREP DET NOUN TO VERB ADJ NOUN 2 1 1
I ) PRON )	= 1.000	(I) * 1.000	* = 1.617827e-07 PRON VERB PREP DET NOUN TO VERB ADJ NOUN 2 1 1

11 1 PRON VERB PREP DET NOUN TO VERB ADJ NOUN 2 1 1 \* = 1.617827e-07

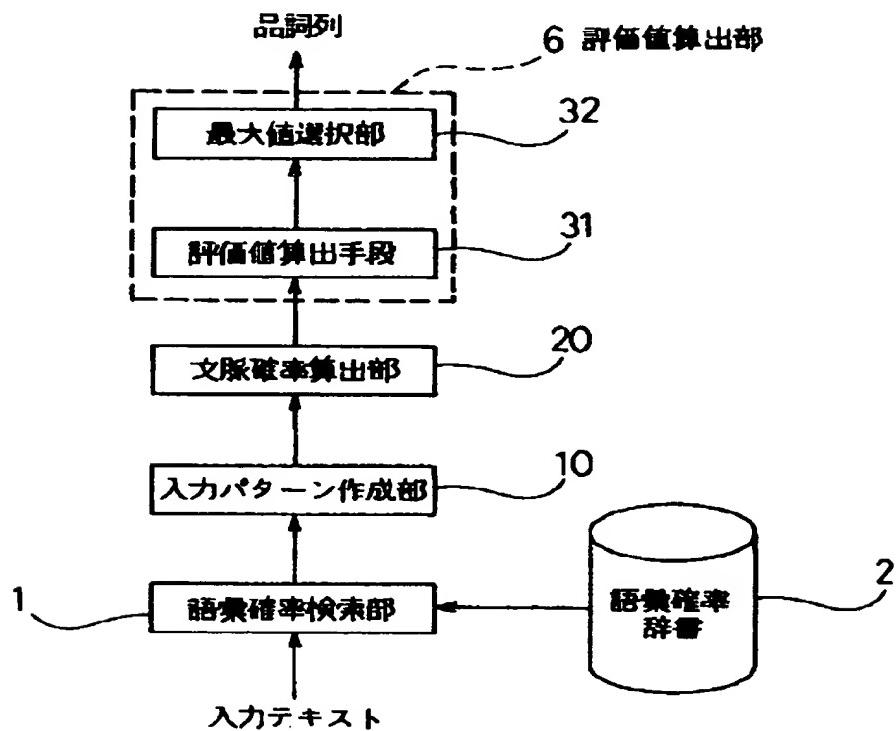
11 1 PRON VERB PREP DET NOUN TO VERB ADJ NOUN 2 1 1

11 1 PRON VERB PREP DET NOUN TO VERB ADJ NOUN 2 1 1

11 1 PRON VERB PREP DET NOUN TO VERB ADJ NOUN 2 1 1

【図16】

## 本発明の他の実施例を示す構成ブロック図



(図17)

### 評価値算出の例を示す図

$$v_n = 1$$

$$m = 10$$

$\text{P}_n = \frac{1}{n} \sum_{i=1}^n P_i$

$\eta_{\text{H}_2} = 6$

$\eta_{\text{H}_2} = 7$

$\text{Vn} = 5$

$\eta_{\text{in}} = 4$

〔图18〕

### 評価値算出の例を示す図

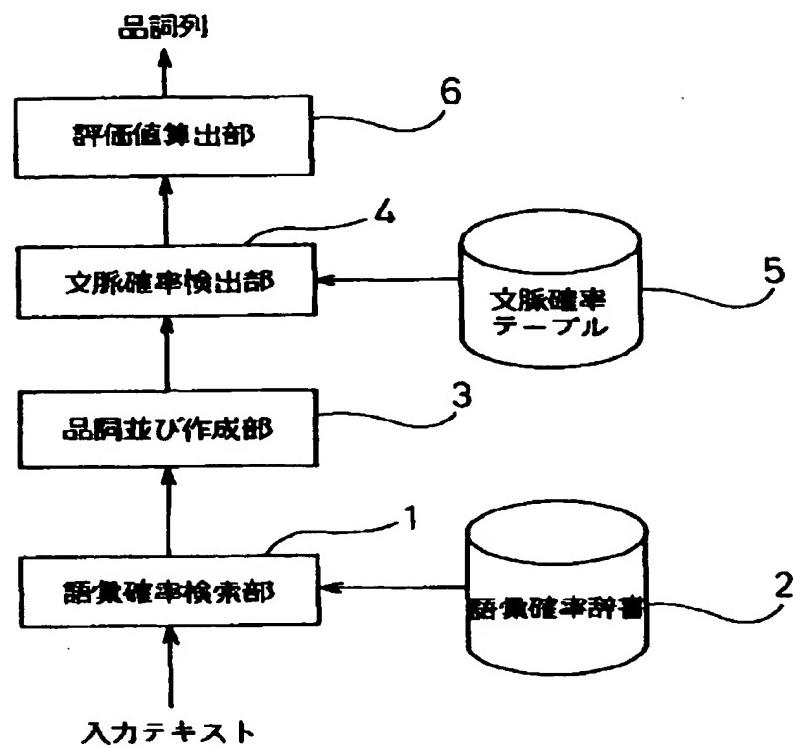
$\text{Vn} = 3$

$m = 1$

$W_{\text{in}} = 0$

【図19】

## 従来装置の構成例を示すブロック図



【図20】

語彙確率辞書の内部構成例を示す図

単語	品詞の種類	確率
I	NOUN	0.00
	PRON	1.00
love	NOUN	0.35
	VERB	0.65
her	PRON	1.00
very	ADV	0.76
	ADJ	0.24
much	ADJ	0.34
	NOUN	0.15
	ADV	0.51
	Z	1.00

【図22】

## 品詞の分類を示す図

品詞番号	品詞記号	品詞名
0	ADV	副詞
1	NOT	NOT
2	ADVP	副詞小辞
3	DET	限定詞
4	N	名詞
5	PRON	代名詞
6	PPRON	人称代名詞
7	NUM	数詞
8	ADJ	形容詞
9	V	動詞
10	AUX	助動詞
11	BE	BE動詞
12	DO	DO
13	HAVE	HAVE
14	PREP	前置詞
15	TO	不定詞を導くTO
16	CJ	接続詞
17	INTJ	間投詞
18	Z	記号
19	#	文頭／文尾